Northumbria
University
NEWCASTLE

A Large-Scale Study of iPhone App Launch Behaviour

Alistair Morrison, Xiaoyu Xiong, Matthew Higgs, Marek Bell, Matthew Chalmers
School of Computing Science, University of Glasgow, UK
alistair.morrison@gmail.com

ABSTRACT
There have been many large-scale investigations of users' mobile app launch behaviour, but all have been conducted on Android, even though recent reports suggest iPhones account for a third of all smartphones in use. We report on the first large-scale analysis of app usage patterns on iPhones. We conduct a reproduction study with a cohort of over 10,000 jailbroken iPhone users, reproducing several studies previously conducted on Android devices. We find some differences, but also significant similarities: e.g. communications apps are the most used on both platforms; similar patterns are apparent of few apps being very popular but there existing a 'long tail' of many apps used by the population; users show similar patterns of 'micro-usage'; almost identical proportions of people use a unique combination of apps. Such similarities add confidence but also specificity about claims of consistency across smartphones. As well as presenting our findings, we discuss issues involved in reproducing studies across platforms.

Author Keywords
App Launches; Usage Patterns; iOS; iPhone; Smartphone Usage; Methodology; Mobile HCI; Large-Scale; Privacy

INTRODUCTION
Mobile devices have become the dominant computing platform of our age, with people spending more time in native mobile apps than either desktops or mobile Web browsing [14,24,35,44]. Smartphones have been described as 'Swiss army knives' [6], with users installing apps to customise devices to particular needs and preferences. Many researchers have studied use of smartphones through patterns of app launches, examining aspects such as usage at different times of day [9,48] or locations [16,49], or creating predictive models to aid in recommendation tasks [26,42].

Two platforms have dominated smartphone use in the last decade: Apple's iOS and Google's Android. As of 2016, it is estimated that these two platforms account for 98% of the market [27]. Android has long had the reputation as the more open platform [11], being based on an open source core and eschewing Apple's app review model—and so often seen as a more appropriate platform for research than the more 'locked-down' iOS [7,18]. Although Apple has made recent efforts to support its platforms' use in research systems through their ResearchKit framework [41], this has been focussed on support for the health domain and collecting of data from external sensors. Apple's tight security model has made it hard for HCI researchers to study the usage of the devices themselves. Consequently, there have been many recent studies of app launch data and general usage on Android—e.g. [5,9,12,17,25,29,40,42,46], several of which we discuss later—but no large-scale study of iOS.

This focus on a single platform presents potential problems in sampling bias when attempting to understand the use of smartphones in general. Although Android has had a dominant market share since 2010 [19], iOS is still thought to account for around a third of all mobile devices in active use [21]. Implications in the literature are often discussed for 'smartphones' and 'mobile devices' more generally, despite effectively excluding a significant proportion of devices in use. Use of such studies as the basis for future design of devices or interaction paradigms implicitly assumes a degree of heterogeneity among users and devices that may be inaccurate.

To overcome the inherent obstacles to recording iPhone user app launch behaviour, we study *jailbroken* users. Since the first iPhone, a subset of users has performed this type of privilege escalation to open up extra functionality on devices [32,43]. By designing our logging for jailbroken devices, we can overcome some of the technical limitations preventing app launch studies, and therefore reproduce Android studies with a large cohort of jailbroken iOS users.

We perform a *reproduction* of several earlier studies. As distinct from *replication,* which would seek to re-perform a previous experiment exactly by matching users, technology and method as closely as possible, a reproduction can look to study similar phenomena, but with different conditions or user populations. Such studies provide an important role, in seeing whether results hold in this fast-changing technical environment, or across varied previously under–studied user populations. Different researchers have argued that mobile HCI is "one of the domains in which reproducing […] studies should be encouraged" [13], highlighting the benefits of being able to continually "take the pulse" [4] of a technical landscape, with each new study in a different context ultimately expanding knowledge of the field.

In this paper, we present a reproduction study with a set of over 10,000 jailbroken iPhone users, reproducing analyses

from a selection of papers [9,13,17,46] that examine use of Android. From studies covering high-level usage statistics to more specialised analysis of user differentiation, we aim to identify where our results show similar patterns to past work on Android, and where they differ. Are there patterns witnessed in both platforms? Are there specific areas where we see large divergences in use?

We also discuss challenges, both uncovered by this attempt to perform comparison of results across platforms, and in large-scale mobile HCI studies more generally.

RELATED WORK

There has been a great deal of quantitative work that investigates use of smartphones by installing logging software on users' devices to collect large amounts of data. Aspects of usage studied have included the times that phones are on or off [38], network activity [48], and times at which phones are in portrait or landscape mode [40]. A large amount of work concentrates on mobile app usage, including the apps people install [23], and usage patterns of app launches [9].

A study by Böhmer et al. [9] was one of the first large-scale investigations into app launch patterns. An example of 'research in the large' [37], the authors wrote software to log app launches, then released this to the public to collect data from 4,100 users. This allowed the study of many aspects of usage, revealing, for example, that communications apps were very heavily used throughout the day, whilst other categories show spikes in usage at particular times.

With a focus on security and user identifiability, Welke et al. [46] analysed the apps that users have launched, to study how easy it is to differentiate between users. The authors found that on average people used 74 different apps, and that over 99% of users had unique usage patterns, even when limiting analysis to the top 60 most used apps.

Ferreira et al. [17] performed a study focussing on the more specific measure of app *micro-usage,* defined as a short interaction session with a mobile application. Clustering app usages by duration, they found a 'natural break' in usage at 15 seconds, with 41.5% of app uses being a shorter duration and thus deemed to be micro-usages. Church et al. [13] also studied this phenomenon, documenting several deployments with varying user demographics and finding a spread of natural break points from 16.6 to 22.5 seconds, representing between 53% and 56% of all usages analysed.

Despite the variety of approaches in studying app launch data, a common trait among all of the above studies is that they were performed on Android. As detailed below, it is a greater challenge to record app launch data on iOS, and Apple places greater restrictions on the functionality of apps it will allow on its Store. Consequently, there have been few attempts to perform analyses of app launches on the iOS platform, and those that do exist involved direct deployment to small numbers of users. For example, in studying data connectivity from the viewpoint of sustainability, Lord et al. recruited 13 local participants and installed logging software on their iOS devices, recording data usage, battery levels and the foreground app, although no details of this logging software are provided [30]. One of the data sets studied by Church et al. had 7 iOS users for whom app launches were tracked for an average of 18 days [13]. McMillan et al. [31] also recorded app launches from iOS devices, alongside screen recordings of the devices. All of these studies were small-scale local deployments, and so did not discuss broad trends in app launch data.

iOS is a completely different platform to Android, with a different underlying OS. Each has a separate 'app store' offering overlapping but different sets of apps. Some fundamentals (such as the basics of homescreens full of app icons, multi-touch gestural input and built-in sensors) are generally shared across the platforms, yet other UI paradigms are distinct to each, such as Android's homescreen widgets. Studies have also identified differences in the demographics of users of the platforms. For example the average iPhone user has been reported to have a higher level of educational attainment and higher average income, Android users are more likely to have technical-related jobs [20], Android devices are more popular in developing countries [34] and iOS users are more likely to buy apps [22], although notably, such reports do not seem to suggest large differences in age or gender distributions. Given this collection of differences, we cannot assume that findings on usage gathered solely from one OS would apply to all smartphone users in general, or *which* recorded aspects of behaviour from past studies might be affected by such factors and show different results on iOS.

This highlights a general challenge in this area of research: reproduction. Each of the studies previously mentioned was conducted with a specific set of users, and results should be read with implications of sampling methods in mind. Small deployments might only recruit local users—whose behaviour may not represent a global population. Even larger scale trials, run through public release of apps that users may choose to install, necessarily recruit self-selecting participants. Issues such as these are at the core of why reproduction is a valuable practice, to widen the overall picture through the undertaking and inter–relation of many studies.

Reproduction in Mobile HCI

Drummond argues for a distinction between replicability and reproducibility in science, with the crux being that replication repeats an experiment avoiding any change from the original [15]. In the domain of Mobile HCI, replication has been interpreted [4] as seeking to validate findings from an original study by re-performing each step exactly as before—from recruitment, to the technologies under examination, to analysis and hopefully arriving at the same results. More common in lab-based environments or physical sciences, this is a significant challenge in the field of Mobile HCI; hardware and software are advancing quickly, and user behaviours are co-evolving with technology, adopting different practices and developing different understandings

of devices through regular usage and exposure to new technologies. Replication of in-the-wild experiments might not be possible if conducted in a world significantly different from the original setting. A reproduction, such as that conducted here, looks to study similar phenomena, but under different conditions or with different user populations. The benefits of reproductions include studying phenomena of interest to assess their generalisation to new users or contexts, and guiding examination of a particular group of interest so as to find how behaviour compares to prior studies.

Several researchers have discussed issues surrounding reproduction of mobile HCI trials [4,10,13]. We identify a number of further challenges which apply to our reproduction across platforms, but also more generally.

Comparison of Logged Data

Many studies have examined app launches on Android. Several frameworks [1, 3] for Android exist that standardise many aspects of data capture and formatting, and public datasets are even available. In writing our own logging tools, and making independent decisions on logging, cleaning and ethical handling of data, it becomes harder to make comparisons between studies. Definitions of how to determine durations of app usage or phone 'sessions' differ [8]; even within the studies covered in this paper, we see a session defined as a period of continued use, broken when the device has 5 [13] or 30 [9] seconds with a locked screen. Data can be processed in different ways, e.g. with different decisions on the filtering of outliers, and what data to discard from a user whose log looks partly corrupted. At an even earlier stage, different data might be recorded due to differing research questions, technical constraints across platforms, or rules enforced by App Stores or ethical policies at institutions [39]. For example, in this study we elect not to log user location beyond broad country-level information, preventing location-specific comparisons.

Comparisons Across Platforms and Time

Although we are comparing results across platforms, we note the extent to which many of these difficulties could also be present in comparing data sets from different time periods. There are difficulties in making comparisons when there were differences in the specific set of apps available, but with apps regularly being released or removed from stores, this could also be encountered in analysing data on the same platform from a different time. Apps change categories, and platform or store owners also often alter the categorisation schemes themselves. At a lower level, technical differences in the functionality of devices could change with new OS versions, and API changes on either platform could lead to new limits in what can be measured or how data is logged, leading to reproduction of studies becoming incrementally harder as time progresses.

Timeliness Challenges

Papers providing high-level statistics on usage can be very informative, and often of great interest; at the time of writing, the study of [9] has been cited over 400 times. However, such a publication is a 'snapshot', documenting observations at a particular moment. In our fast paced field, people and use are likely to change quickly, and results will soon no longer accurately describe the current world.

If the goal were to keep an audience informed with information relevant to the current state of affairs, one option would be to publish such papers at regular intervals. However, would a researcher want to keep redoing the same analysis? There has been much discussion as to the role of replicating research findings in HCI [47]. It is not clear to what extent work presenting purely 'catch up' quantitative findings forms part of that discussion, and whether such a paper would be accepted for an HCI venue unless also testing some novel phenomenon or new user groups. Unless contextualising results with more qualitative data, it is arguable whether such studies even fall within the role of HCI research or academia in general, and might simply be the domain of marketing or analytics companies. Another option might be to change the mode of experimentation from separate trials generating their own discrete data sets, interspersed with phases of comparison and commensuration, to one ongoing experiment, collectively organised and continually analysed by the research community.

We suggest that future comparative research and meta-analysis would be aided greatly by standardised recruitment procedures, mature ethical policies that balance the protection of personal data with the value of insights obtained from cross study collaborations, and standardised ways in which usage data is not only logged, but cleaned, processed and measured. Toolkits/APIs such as AWARE [3] are a good starting point, in technical terms, but we may also learn from the similar challenges seen when integrating databases or software systems, and so develop data meta-models that express semantics in utile ways.

We also note that the majority of study findings in this area centre on relatively simple descriptive statistics and charts, e.g. means of app usage duration and hourly charts of key app usage levels. There is, of course, a plethora of data analysis and visualisation techniques available today. We note our own work on enriching the analysis of app usage by incorporating 'side information' in the form of rich text data (from app stores) describing such apps [45] and probabilistic discrete state models of app usage patterns [2]. Our focus here is on reproduction, and so in this paper we apply the same models as prior studies, but we suggest that there is a research gap here that future work should fill.

A REPRODUCTION STUDY ON JAILBROKEN IPHONES

User Subpopulation

In this paper we present AppTracker – a logging platform for large-scale collection of iOS app launch data. Tracking app launches on iOS is a technical challenge. There has never been a public API providing access to launch histories or apps in use. Apple also places restrictions on apps running in the background, which is necessary in order to

constantly monitor certain aspects of phone state. Even where technical workarounds can be found to overcome backgrounding restrictions, Apple will still enforce this rule during the app review process [39], and prevent such software from being released through the official App Store. As such, AppTracker was released via an unofficial third-party repository for jailbroken iOS devices (see [32] for details), and our data is gathered solely from such devices.

It is uncertain how the average jailbroken user would behave as compared to the average non-jailbroken user, and the extent therefore to which our data set is representative of iOS users in general. Past work has suggested that on average jailbroken users were older, skewed more towards males and were perhaps more technically literate [32]. It is unknown, however, whether those demographics hold true generally or were particular to users of the specific game under investigation. People might jailbreak to gain extra functionality, so might be early adopters, or particular phone enthusiasts [36]. If we consider the users from the various Android studies as being more representative of a more general population, we might suspect that our users would use their devices for longer, and use a larger collection of apps. They might have a less 'mainstream' selection of apps, with jailbroken users also able to install many more apps not available on Apple's App Store [32], or might experiment with installing new apps more often. Consequently there might also be less overlap among users in used apps. In a study of phone 'sessions', van Berkel et al. discuss how a phone usage session "can range from brief tasks confined within a certain application to overarching tasks spanning multiple applications and services" [8]. We might also assume that our jailbroken cohort are 'power users', who might show greater tendencies towards the latter end of that spectrum than more general users.

Like many other similar studies [9,40,46], our data is recorded from a publicly released app, whose users are entirely self-selecting and would presumably only download an app and keep it installed if they were interested in monitoring their own device usage. These practices could also lead to a bias that recruits users with specific interests or behaviours. We acknowledge these potential sampling biases, but consider that our approach represents the best available way to perform large-scale analysis of iOS app launch data, and to effect a comparison with prior Android–focused studies. Beyond this we believe that the cohort of jailbroken iPhone users in itself offers an interesting group to study.

Reproduced Analyses

This paper presents a reproduction of the analyses described above from Böhmer et al. [9], Ferreira et al. [17], Church et al. [13] and Welke et al. [46]. As this is the first such large-scale study considering iOS devices, we sought to reproduce a spread of studies. Our high level aims are to perform an initial investigation of jailbroken iOS users, discovering the areas where our results show similar patterns to past work on Android, and where they differ.

We believe that reproducing the analysis of [9] is valuable in providing high-level stats that offer a good general overview of the behaviour of our user set. This study covers many aspects such as session lengths, usage across time of day and app categories, that encapsulate basic day-to-day device usage, but which are hitherto unexplored on iOS devices. The micro-usage studies of [17] and [13] allow further drilling into such summary statistics; if different average app use durations are observed, are they explicable by varying degrees of this phenomenon of micro-usage? These studies are also good candidates for reproduction as they already provide a spread of results across a number of user groups. This affords an opportunity to investigate whether our results fall close to or within that spread, or stand far apart, providing a further indication of the similarity of our cohort's usage's to the existing literature. Finally, the study of [46] has important implications in user identification based on used apps. iOS has a different set of apps available to users, and jailbroken users have access to more still, so this seemed a good study to investigate whether statistics from identifiability techniques based on unique apps used would generalise to our data set. This issue also has particular significance for HCI research communities in general, where anonymisation might be considered a prerequisite to sharing datasets. Our aim is that, together, these studies provide a fair cross-section of the Android studies previously conducted, and are therefore suitable as a first large-scale look at iOS in general.

A meta-concern for us across all studies was the extent to which they would be generalisable at all, given issues such as technical differences and sampling processes. In our discussion, we document experiences and challenges in our attempt to reproduce these analyses from earlier studies.

To help present our findings, we structure them around 3 hypotheses. Firstly, the majority of the most popular apps identified by past studies [13,17] are also available on iOS. Strong trends have also been seen in categories, such as Communications being very dominant [9]. Although we think that jailbroken users might trial many new apps, these would not necessarily be at the expense of the popular apps used in general. As there have been so few large-scale studies of iOS use, we formed our hypotheses with an open-minded approach; we ran our study as a first exploration of whether differences would emerge, and often had no reason to suspect that our observed data would differ from the Android results in the literature, so formulated our first hypothesis as such.

H1: There will be very similar usage across platforms in broad trends in device usage, such as: popular categories of apps, times of day the phone is used, times of day particular categories are used

While H1 looks at *when* phones are used and *which* specific apps in use, H2 looks more at *how much* usage. Concentrating more on differences in our user population as described in the previous section, our second hypothesis covers the

idea that our users might be more enthusiastic phone users, or show 'power user' characteristics.

H2: Our sample are more enthusiastic phone users: they will install more apps, use their devices for longer, and tend to use more apps in a single session

Finally, in H3 we propose that different sets of available apps across platforms, and our users potentially being motivated to jailbreak in order to gain access to greater customisation and a wider variety of apps, mean that there will be lower degrees of overlap between users w.r.t. used apps.

H3: The proportion of 'anonymous' users and patterns of user proximity [46] will be lower in jailbroken iOS

APPTRACKER

Our data is collected via AppTracker, an app for jailbroken iOS devices consisting of a background logging framework that captures information on device use, and a foreground UI that displays charts and statistics on app use durations.

AppTracker operates by running in the background to sample high-level information on the state of the device. It records timestamped logs of every time an app is opened (whether from the home screen, multitasker, a notification, or any other method) or closed on the device. It also tracks every time the device is locked or unlocked (whether through pressing the sleep/wake button or through auto-lock after a period of inactivity). Through these measurements, we can determine how long a user has an app active in the foreground, and these timestamped logs allow us to collect similar data to, and therefore reproduce, all our selected studies. No information is recorded on activity within individual apps, or activity over networks.

Screenshots from the AppTracker user interface are shown in Figure 1. In order to encourage users to download AppTracker and keep it installed, it displays a rich set of charts and statistics on the user's app usage. Users can view lists of the most-used apps, filter by time period, or view detailed information on usage of an individual app.

AppTracker regularly uploads user app launch logs to our servers, together with a timestamp, device type, device identifier, and timezone. On first launch of the app, a terms and conditions page explains the data that will be logged and requires explicit consent before the application can be used. In terms of McMillan et al.'s ethical framework for large-scale Mobile HCI research projects [33], AppTracker would be classed as logging data that would be expected given the apparent functionality of the app, and we consider it an open question how personally identifiable a user might be from their app launches.

Collected Data

AppTracker has been downloaded over 40,000 times as of September 2017. The app can run on the iPhone, iPod Touch and iPad. Users have supplied 28 million app launch events. Except where explicitly stated otherwise, for the purposes of this paper, we consider data gathered from

Figure 1. The AppTracker interface shows summary statistics and charts to inform the user about their device usage

10,338 iPhone users who used AppTracker between 8th August 2013 to 30th January 2017. Most come from Europe (3332), North America (3049) or Asia (2554), cumulatively accounting for 86% of all users. During this time frame of around three and a half years, launches of 45,803 unique apps were recorded.

In the following sections, app 'categories' are discussed. For most apps, this is a developer-assigned categorisation, which we retrieved from the App Store. Where this was not possible (for example, Apple's first party apps that are built into iOS are not on the Store), we manually assigned apps to categories. We also removed the launcher apps, such as the 'Springboard' main menu on iOS from the results.

In the results sections, we refer to an unbroken period that an app spends in the foreground on the device as one *app usage*. By our definition, a usage ends when the user returns to the home screen, or brings a different app to the foreground. When the screen locks with an app open, we also consider this to be a usage ending, and if the screen unlocks to reveal the same app still in the foreground, we consider this a new usage beginning. Also discussed is a *usage session:* a sequence of one or more app usages. This concept is measured in different ways in the literature [5,8,9]. Here, we follow Böhmer et al., who mark a session end when a device's screen has been locked for 30 seconds.

FINDINGS

Descriptive Statistics

We studied general patterns in our app usage data, aiming to recreate analysis by Böhmer et al. on data from Android devices [9] looking at categories, times of day of app launches and the chaining of usages into app sessions. Table 1 shows results from our data, aggregating apps by category. For each category we list example apps, the total number of app launches recorded during our study and the mean length of an app usage. We can see that Communica-

tion is by far the most used type of app, at 4.8M total launches. Social Networking is the only other category to top a million launches in our study. There is a large difference in the most and least used categories. The longest category we see by some distance is Health & Fitness. We found that this was being caused by one sleep-tracking app, which although little used, would have sessions of 8 or 9 hours in length, pulling up this average. Without this app, the mean duration for Health & Fitness would be 52.6s.

Category	# App usages	Avg. duration	Example Apps
Weather	34012	28.9	Weather, Yahoo Weather, the Weather Channel
Settings	328515	34.5	Settings, SBSettings
Food and Drink	1523	44.5	Starbucks
Finance	16131	53.5	Passbook, Stocks, Bank of America
Productivity	192325	55.8	Evernote, Notes, Reminders, Google Drive, Dropbox
Communication	4784608	60.5	Skype, Mail, Messages, Phone, Facebook messenger, WhatsApp, Google hangouts
Travel	19221	63.8	Gasbuddy, Foursquare, Yelp
Utilities	734235	66.4	AppStore, Calculator, Compass, Clock, Calculator
Business	4647	67.3	Adobe Reader, Facebook Page Manger
Sports	7509	93.3	ESPN ScoreCenter
Photo and Video	869197	95.0	Camera, Instagram, Youtube, Flickr, Photos
Music	304673	100.0	Music, Pandora, Spotify
Lifestyle	49746	102.8	Amazon, eBay, Groupon, Jolly
Reference	10674	106.9	Google Translate, Dictionary, Bible
Social Networking	1224070	115.1	Facebook, Twitter, GooglePlus, LinkedIn, WeChat, Pinterest
Navigation	130850	182.0	Google Map, Apple Map, Navigon, Waze
Browser	703486	136.6	Safari, Chrome, Puffin Browser, Opera
News	56868	210.8	CNN, Alien Blue (Reddit), Feedly
Games	159720	293.0	Fruit Ninja, Flappy Bird, Clash of Clans
Entertainment	95317	296.8	Netflix, imdb, podcasts, Fandango
Books	10778	315.1	Kindle, iBooks
Health and Fitness	13981	2078.9	RunKeeperPro, Moves

Table 1. Number of app usages and the average app usage duration (in seconds) for each category

Comparing to Android, we see that similar categories are most used across platforms, with Communications very dominant in both, lending support to H1. Yet we also note that our usage appears higher in general, supporting H2. The average app use duration in our data is 88.6 seconds – around 17 seconds (24%) higher. The rankings of categories by average duration are different across the platforms, yet our users spend slightly longer in apps for most categories (e.g. Communication 60.5 seconds vs. 46.9).

App Usage By Hour of Day

Figure 2 presents the number of apps launched per hour of a day. It has a very similar pattern to Android, again supporting H1. Most app usages happen in afternoon and evenings. However, our data shows more diversity over the day, with the busiest period seeing around 9 times more launches than the quietest. On Android, this is a 6 times difference.

In Figure 3, again inspired by a table in [9], we show the hourly change in the relative usage of the app categories in terms of number of app launches. Each column is a distribution over the app categories used in a particular hour, and each row is then coloured light yellow to green to reflect when each app category has its highest relative prominence in usage. We can see that Communication apps dominate, and are the most likely to be used every hour of the day, particularly in the afternoon and evening with a chance higher than 50%. Such is the dominance of Communication, that many other categories get their relative peaks (shown by darker colouring) during night hours, when Communication relents somewhat. Various other apps show relative prominence at other periods, such as Weather from 6am to 8am and Sports later in the day.

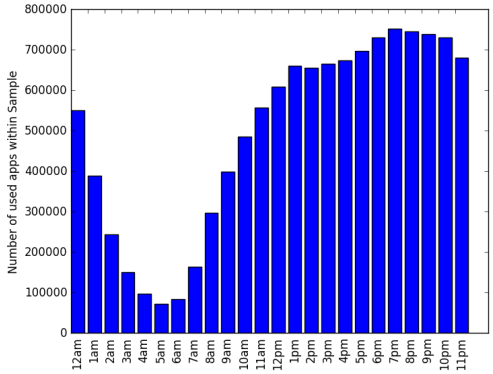

Figure 2. Number of app launches by hour of day.

Looking at results from Android, we can see some clear similarities, including the consistent high proportion of Communication, rising to over 50% of use between 11am and 11pm. The colouring of most category rows are very similar, including Sports in the afternoon and evenings and Games starting in the evenings and throughout the night. Again, this evidence supports H1. However, as not all categories match between the two platforms, it is difficult to perform a complete comparison.

Characterising App Usage Sessions

Following [9], we consider a 'usage session' to be all activity performed in a continuous usage of a device, terminating when the screen has been locked for 30 seconds. There are a total of 3,750,305 such sessions in our data. Figure 4 shows the distribution of apps used in sessions. In general, most sessions are short, with 83.5% containing 4 apps or fewer. Reported Android findings show a similar long tail distribution, but in general sessions in our data seem longer. A majority of Android sessions (68.2%) contain a single app, whereas in our data this is only 38%. It is reported that 94.3% of Android sessions have 3 apps or fewer, but here this is only 76.7%. This is support for H2.

Understanding Micro-Usage

App *micro-usage* is a concept introduced by Ferreira et al. [17], defined as short bursts of interaction with apps. They partitioned their app usage data into 2 clusters, looking for a 'natural break' point to separate their data into micro-usages and non-micro-usages. They found this break at approximately 15 seconds, with 41.5% probability of micro-usage. Church et al. reproduced this on 3 data sets, finding 3 different break points of 16.6s, 22.5s and 21.5s, representing 53%, 56% and 55% of all usages analysed [13]. We similarly ran k-means (k=2) to determine the natural break in our jailbroken iPhone data, finding it at 21.4 seconds, with 43.6% of our app usages thereby being classed as micro-usage. Interestingly, the breakpoint we find is most sim-

	12am	1am	2am	3am	4am	5am	6am	7am	8am	9am	10am	11am	12pm	1pm	2pm	3pm	4pm	5pm	6pm	7pm	8pm	9pm	10pm	11pm
Books	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.2%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%
Browser	7.9%	8.4%	9.1%	9.5%	9.9%	10.1%	9.5%	8.6%	8.3%	8.0%	7.5%	7.1%	6.7%	6.8%	6.8%	6.8%	6.8%	6.6%	6.6%	6.7%	6.9%	7.0%	7.0%	7.3%
Business	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Communication	44.4%	42.0%	39.3%	37.4%	36.2%	35.4%	36.6%	38.5%	42.3%	46.0%	49.6%	51.4%	52.2%	52.3%	51.9%	51.8%	51.6%	52.0%	52.0%	51.4%	50.5%	49.8%	48.8%	46.8%
Entertainment	1.2%	1.3%	1.3%	1.3%	1.3%	1.3%	1.1%	1.2%	1.1%	1.0%	0.8%	0.8%	0.8%	0.8%	0.8%	0.9%	0.9%	0.9%	1.0%	1.0%	1.0%	1.0%	1.0%	1.2%
Finance	0.1%	0.1%	0.1%	0.1%	0.2%	0.2%	0.3%	0.2%	0.3%	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%	0.1%	0.1%	0.1%	0.1%	0.1%
Food and Drink	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Games	1.7%	1.8%	1.9%	1.7%	1.7%	1.8%	1.7%	1.8%	1.8%	1.8%	1.6%	1.6%	1.5%	1.5%	1.5%	1.6%	1.7%	1.6%	1.5%	1.6%	1.7%	1.7%	1.7%	1.7%
Health and Fitness	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%	0.4%	0.5%	0.3%	0.2%	0.2%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.2%
Lifestyle	0.6%	0.6%	0.6%	0.6%	0.7%	0.7%	0.6%	0.5%	0.5%	0.5%	0.5%	0.5%	0.5%	0.5%	0.5%	0.5%	0.5%	0.5%	0.5%	0.5%	0.5%	0.5%	0.5%	0.5%
Music	2.6%	2.6%	3.0%	3.4%	3.3%	3.4%	3.7%	4.1%	4.9%	3.9%	3.3%	3.0%	3.0%	3.1%	3.2%	3.3%	3.3%	3.3%	3.4%	3.2%	2.9%	2.7%	2.6%	2.6%
Navigation	0.8%	0.7%	0.7%	0.8%	0.9%	1.0%	1.1%	1.1%	1.5%	1.5%	1.4%	1.3%	1.5%	1.5%	1.6%	1.6%	1.6%	1.6%	1.8%	1.7%	1.4%	1.1%	0.9%	0.8%
News	0.7%	0.7%	0.8%	0.9%	0.8%	0.8%	0.8%	0.8%	0.7%	0.7%	0.6%	0.5%	0.6%	0.6%	0.5%	0.5%	0.5%	0.5%	0.5%	0.5%	0.5%	0.5%	0.6%	0.6%
Photo and Video	10.0%	10.0%	10.2%	10.5%	9.9%	8.8%	7.6%	7.4%	7.7%	7.6%	7.6%	7.9%	8.1%	8.2%	8.4%	8.5%	8.8%	9.1%	8.9%	9.0%	9.4%	9.6%	9.8%	10.0%
Productivity	1.6%	1.6%	1.6%	1.8%	1.8%	2.1%	2.2%	2.2%	2.4%	2.6%	2.6%	2.4%	2.2%	2.2%	2.1%	2.2%	2.1%	1.9%	1.9%	1.8%	1.8%	1.7%	1.6%	1.6%
Reference	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%
Settings	3.4%	3.8%	4.2%	4.4%	5.0%	4.8%	4.5%	4.1%	3.7%	3.9%	3.5%	3.3%	3.3%	3.1%	3.2%	3.3%	3.2%	3.2%	3.2%	3.2%	3.3%	3.2%	3.2%	3.3%
Social Networking	15.1%	15.3%	15.0%	14.4%	14.2%	13.5%	12.3%	13.0%	12.4%	11.9%	11.5%	11.6%	11.5%	11.5%	11.5%	11.3%	11.4%	11.3%	11.4%	11.8%	12.6%	13.4%	14.4%	15.1%
Sports	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%
Travel	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.2%	0.2%	0.2%	0.2%	0.2%	0.3%	0.3%	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%	0.1%	0.1%
Utilities	8.7%	10.0%	11.0%	12.1%	13.2%	14.7%	16.2%	14.0%	10.6%	9.0%	7.9%	7.3%	6.9%	6.7%	6.6%	6.6%	6.6%	6.4%	6.4%	6.3%	6.5%	6.6%	6.9%	7.6%
Weather	0.3%	0.3%	0.3%	0.3%	0.3%	0.5%	0.9%	1.1%	1.0%	0.6%	0.5%	0.4%	0.4%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%

Figure 3. Hourly relative amount of app launches by category. Each column is a distribution over categories of app launches for an hour. Colours normalise by row: dark showing when each category sees its maximum relative percentage of launches.

ilar to two of Church et al.'s datasets, but the proportion of usages we classify as micro-usages is more in keeping with Ferreira et al.'s original study. Our data is not an outlier among the previous studies, and this offers support for H1.

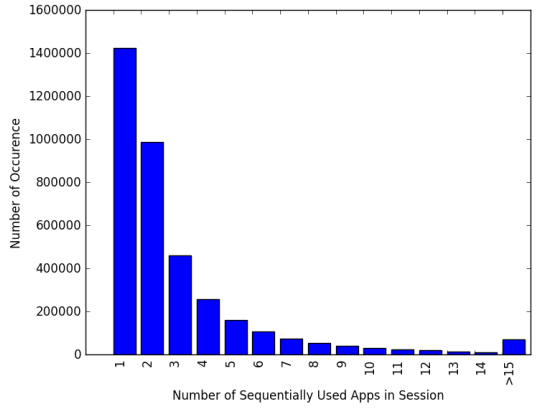

Figure 4. Number of all used apps in a session. Sessions longer than 15 apps were aggregated since the graph flattens out.

We further investigated micro-usage for our top ten most frequently used apps, with Figure 5 showing the probability distribution of the usage duration. In Figure 6, we look at proportion of micro-usages, comparing to Ferreira et al.'s top 10. As with the Android findings, different apps exhibit different micro-usage patterns, with for example only 31.4% micro-usages of Facebook, but 61.6% of email. Although the exact proportion of micro-usages for each app is different between the two studies, we found the general trends of this chart to be very similar. For example, in both studies, Facebook, browser, App Store and WhatsApp all have quite high proportions of micro-usage, with calendar, alarm clock, and email all lower. These close matches in patterns of use are further support for H1.

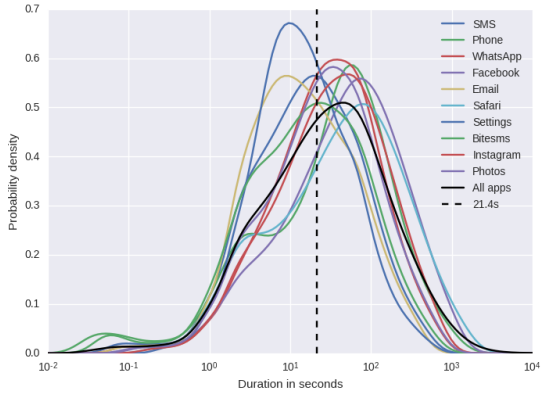

Figure 5. Distribution of app use duration for the top 10 apps overall. The dashed line shows the break point at 21.4 seconds.

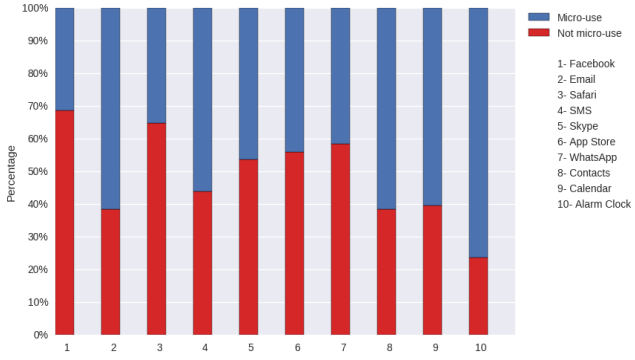

Figure 6. Micro-usage distribution for the ten most popular apps as identified by [17].

Identifying Users From App Signatures

In this section, we reproduce Welke et al.'s [46], study of the ability to differentiate users from their app usage. This study presents more of a challenge in preparing a set of our data with which we can make fair comparisons to the original results. The data consists of binary vectors for each user, indicating whether a particular app was ever launched during the period of observation. As such we thought it important to study data recorded over a similar length of time. Excluding users for whom we had fewer than 5 days' usage, we created a set where users were logged for an average of 46.1 days, which was close to the 48.6 of [46]. This set has 3,574 users who collectively used 43,876 different iPhone apps. This is considerably smaller-scale than the original, which had 46,726 users and 146,532 apps.

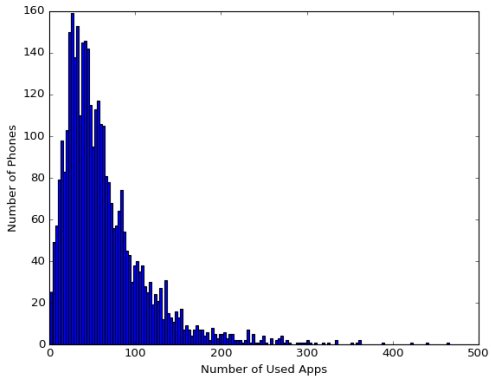

Figure 7. Histogram of the number of used apps per phone. The average is 64.0.

The average number of apps per user is 64.0 (st dev. 52.5) (Figure 7). Although this is lower than the 74.37 (st dev. 44.16) observed in the Android study, the shape of the distribution is very similar. This is evidence against H2.

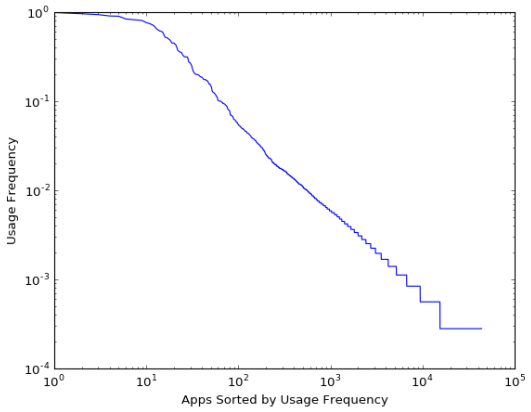

Figure 8. Usage frequencies of top 43,876 apps (log-log plot).

Although based on a smaller number of apps, we found a very similar pattern of usage. Only a few apps are used by a large number of users. Figure 8 shows the plot of the usage frequencies of all the 43,876 iPhone apps, ranked in the x-axis by usage frequency. AppTracker was obviously used by everybody, with the second most popular app used by 95.9% of users. The corresponding figure from the Android study is 95.4%. The 15th most popular app was used on 60.0% of all phones, after which popularity drops quite sharply, with only 15.0% of all users using the 50th most frequent app, and 5.4% using the 100th. The Android study showed a slightly more gradual drop-off, with the 50th 22.8% and the 100th 10.44%. These results support H1.

Welke et al. use the term *app signature* to refer to the set of apps launched at least once by the user, taking the top 500 most used apps, and representing a user as a 500D binary vector with the i^{th} entry being 1 if the user used the i^{th} most frequent app, or else 0. Due to the different characteristics of our data, we elected to use the top 300 apps. This seemed a good choice, as the 300th most frequent app was used by 1.7% of all our users, closely matching Welke et al.'s 500th most used app being used by 1.69% of users. The similarity between 2 users can be quantified by the Hamming distance between signatures – the number of apps exclusively used by either user. 0 Hamming distance between 2 users implies they have used the same set of apps, and are considered in Welke et al.'s discussion as 'anonymous'. The 'uniqueness' of a user can also be measured by distance to nearest neighbour. Out of 3574 users in our data set, only 0.33% (12) are anonymous, with 99.67% of users uniquely identifiable by app signatures – an identical proportion of anonymity as the Android study (153 of 46,726 users; 0.33%). This strongly refutes our hypothesis H3.

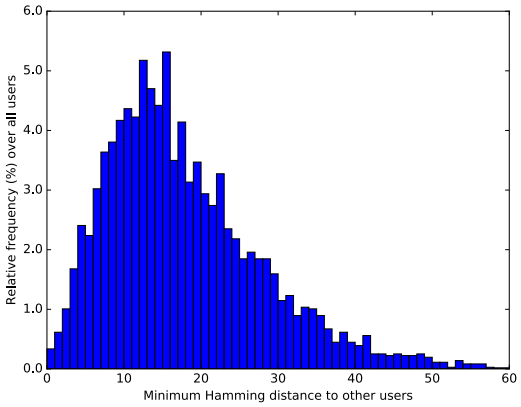

Figure 9. Distribution of the Hamming distance to the closest user based on top 300 apps. The average is 17.4.

Figure 9 shows the distribution of uniqueness; that is, for each user, the Hamming distance to the closest other user. As explained, the data we prepared has different characteristics than that of Welke et al. and this is reflected here. Our results show a lower average minimum distance of 17.4 compared to 25.9 in the Android data. 94.0% of our users have Hamming distances of at least 5; in the Android study, 95% of users had a Hamming distance of at least 10. However the similar shape of Figure 9 to its equivalent in the original is further evidence against H3.

Figure 10 shows the same measure, but based on the top 60 apps. In this case, the number of anonymous users increases to 21 and the average distance drops to 6.8. In contrast to

the previous example, the Android study in this case shows a *lower* average minimum Hamming distance of 4.9. As we are examining *minimum* Hamming distances, the larger number of users in the Android study increases the likelihood of pairs of users being similar and this figure decreasing. In the Figure 10 example, the Android user sample was still larger, but we were comparing vectors of size 300 vs 500, these larger vectors raising mean Hamming distances.

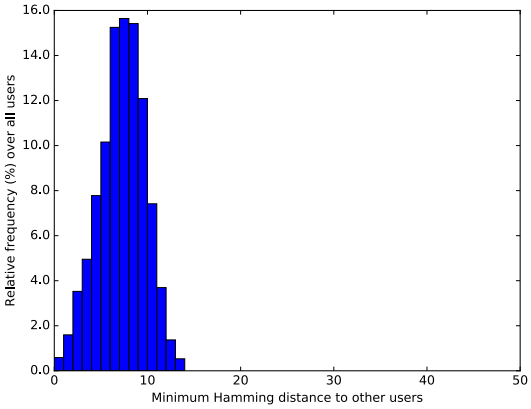

Figure 10. Distribution of the Hamming distance to the closest user's app signature based on top 60 apps. The average is 6.8.

In general, it can be seen that this study was difficult to reproduce. In contrast to our previous examples, the results are very tied to the sample sizes, and despite our efforts to process test sets with comparable characteristics, some differences in the data will remain difficult to reconcile until more data is collected. However, from what we have been able to show, the evidence does not support H3.

DISCUSSION

Comparison of Results with Original Studies

Evaluating H1, that general patterns of device and app usage would be similar across platforms in terms of popular apps, categories, and times of day, we see a lot of evidence in support. Most of the distributions and rankings computed show broadly similar patterns, particularly among the strongest trends. For example, communication was by far the most popular category of app on both platforms, seen to be the most likely type of app to be used at all times of day or night on both. Most-used app lists were comparable, and broadly similar patterns in usage could be seen across many categories of app. Differences that do exist may be due to specific apps or categorisations between platforms. Micro-usage behaviour among our users seemed to correspond to that reported on Android, and our results fall within the variety already seen among multiple Android studies [13]. Examining specific apps in detail showed very similar micro-usage trends between platforms. Therefore, as well as individual apps showing large overlap, *how* these apps are used also seems similar across the user groups.

Our second hypothesis H2 concerned itself with our jailbroken cohort showing more intense usage of their devices, or 'power user' characteristics. We see more mixed findings here. There is evidence in support, in terms of our users using their devices for longer, with individual app usages being around 24% higher than Android (a trend seen overall and across most categories). As mentioned above, it does not look as if these longer average app usages we observe are explicable by a smaller proportion of micro-usages. We also saw more apps used in sessions; indeed, Android had the majority of the sessions showing a single app usage, whereas single-app sessions accounted for only 38% of those seen on our data. This could be evidence of our jailbroken iOS cohort being more likely to show a more sophisticated use of apps in combinations, in line with H2. However, evidence against H2 is also apparent in that we see that, generally speaking, a typical user in our study actually uses a smaller set of apps than a typical Android user.

Although the user differentiation study proved more of a challenge to compare across platforms, we can again see similar usage patterns, with a small set of apps being used by large percentages of users, and app popularity fast dropping and showing long tails. On both platforms, people use quite distinguishable sets of apps, with identical small proportions being classed as 'anonymous', and similar usage frequencies and distributions of proximity to neighbours. This evidence refutes H3, based on an assumption that our users would have access to a large collection of apps, which they would regularly install and experiment with, and consequently show less overlap and lower anonymity. Instead, the result does provide support for Welke et al,'s original result generalising more broadly across smartphones.

Although some distinct usage is apparent between the different user groups, we cannot provide conclusive explanations for the differences observed. It could be the case that people with different goals or attitudes are attracted to each platform; that different user experiences are offered that might encourage different behaviours; or specific properties of the hardware such as physical size, weight or battery life might cause different levels of enthusiasm for longer usage sessions. Similarly, our users showed more apps per session on average, which could be due to jailbroken users showing mastery of sets of apps in combination, or equally due to the implementation of app switching interfaces in iOS. In general, although we have gathered certain pieces of evidence in support of or against our hypotheses, we do not claim to provide a full picture of user motivation or reasons for differences observed. Although our quantitative study has revealed how users behave, further studies (and perhaps the use of qualitative or mixed methods) are needed to explain our observed differences with more certainty.

Challenges Encountered in Comparing To Past Studies

Comparing across platforms seemed feasible for broad concepts such as durations or app counts. However, we saw that comparing on specific apps could be more challenging. For example, a few launches of a sleep-tracking app skewed the average duration of use for an entire category disproportionately, given its comparative lack of use.

We encountered a few issues with categories in our analyses. For example, we were considering data, both from our logs and past publications, that were often years old and several apps had changed categories (for example, Twitter [28]). Of our 22 categories, there are 15 exact matches to category names quoted in Böhmer et al.'s paper. Some seem to have different labels but map quite well (for example, Tools vs. Utilities). For others, Android appears to join sets of apps that iOS keeps distinct. For example, there are separate Travel and Navigation categories on iOS, but on Android, Travel alone seems to cover both. Such subtleties in categorisation have an effect on comparisons. For example, in Figure 3 we see high points in relative use for Navigation around working commute times. In Android data, we do not see overall category peaks at these times, possibly because other non-Navigation Travel apps drown this out.

Our final set of results, recreating the analysis of Welke et al. on user differentiation, caused the greatest difficulty. This was a study where it seemed that the size of the user sample would not just have an effect on the validity of the results, but would have a substantial bearing on the actual results themselves. As the study measures 'uniqueness' of users via the distance to each user's single nearest neighbour, it seems necessarily the case that increasing the size of N will yield closer nearest neighbours on average.

As the Android study had more users than our own, we could not match sample size. Additionally, the Android study had only considered data from users who completed a questionnaire, possibly yielding a more 'committed' cohort. These users were filtered to those with at least 1 day of use, giving a set who had used the app for an average of 48.61 days. When we excluded users with less than 1 day of use, we did not arrive at a set with a similar average number of days' use. Which is the most important criterion to match? We chose the latter and approximated it well by excluding users with fewer than 5 days' use of AppTracker.

Although this selection allowed a comparison to be made, it had the consequence of requiring further sampling decisions later, and we elected to use 300 apps rather than 500 in Hamming distance calculations. Eventually we actually found a similar proportion of users deemed 'anonymous' across the studies, but this comparison comes with many caveats and we were unable to present a 'clean' result.

CONCLUSION

Study reproduction is a valuable practice in HCI. Not every piece of research can consider all groups of users or study a phenomenon over many years, so each study is likely limited to considering a relatively small number of users at one time. For the field to keep learning, it is necessary to determine whether results still hold, among different subpopulations, or as environments evolve. We have attempted to reproduce previous studies on the use of Android, hoping to gain a general sense of whether findings would also generalise to our users, and the extent to which such comparisons were even possible. We collect data from jailbroken iOS users, who form an interesting cohort in itself, but we do not claim our findings would generalise to all iOS users.

We present our findings based around 3 hypotheses. We found evidence to support H1, our hypothesis that users would behave very similarly across platforms among strongly observed trends such as apps for communication being the most used on both platforms, proportionately highest in the afternoon and evening; most apps being used by a small proportion of users, and only few apps used highly frequently; over 99% of users differentiable by their app signatures over a 45 day period, and micro-usage behaviour being within variation spreads already seen [13].

We found a more mixed message in examining H2, where we assessed whether the jailbroken iOS users would be more intense device users and be regular eager experimenters with new apps. Our users did use apps for longer, across most categories, and they combined more apps into their usage sessions, but on average they used a smaller set of apps over 45 days than users in the previous Android study,

Our third hypothesis was quite strongly refuted by our data – there is not a lesser degree of overlap in the used apps of our users, and Welke et al.'s results hold across platforms.

For many trends, therefore, it seems that even though we have recruited users on a different platform, from a specific subpopulation, there is a high degree of similarity in usage, and we can be more confident in claiming that these patterns are consistent across smartphones in general. However, our reproduced studies have uncovered enough distinct behaviour to suggest that there is heterogeneity of use across platforms. Having identified which areas show distinct usage patterns, we also offer a challenge for future work: to further explore these, possibly through qualitative methods, to ascertain why results appear as they do. Moreover, we propose that, in designing for future generations of smartphones, it is not sufficient to constantly draw implications from the same platform-specific subset of users.

We also discuss challenges encountered in our (and other) comparison studies. The two platforms under consideration share enough in terms of fundamental design that we could simply compare broad concepts such as counts of apps or usage durations. More difficulties arose in terms of specific app availability, categorisation, sample matching, data logging intricacies or varying ethical practices across institutions. However, many of these issues are not unique to cross-platform studies, and could arise in any attempts to re-examine previously published results. We propose a number of responses to such challenges, some technical (e.g. developing standardised frameworks for experimental recruitment, and for data representation and processing) and some practical (e.g. more continuous and communal approaches to such experimentation). We hope to contribute, along with others, to such future developments.

ACKNOWLEDGMENTS

This research was supported by EPSRC (EP/J007617/1).

REFERENCES

1. Ionut Andone, Konrad Błaszkiewicz, Mark Eibes, Boris Trendafilov, Christian Montag, and Alexander Markowetz. 2016. Menthal: a framework for mobile data collection and analysis. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct* 624–629. DOI:10.1145/2968219.2971591
2. Oana Andrei, Muffy Calder, Matthew Chalmers, Alistair Morrison, and Mattias Rost. 2016. Probabilistic Formal Analysis of App Usage to Inform Redesign. In E. Ábrahám and M. Huisman (Ed.), *Integrated Formal Methods: 12th International Conference, IFM 2016, Reykjavik, Iceland, June 1-5, 2016, Proceedings*.115–129. doi:10.1007/978-3-319-33693-0
3. AWARE. Open-source Context Instrumentation Framework For Everyone. http://www.awareframework.com
4. Nikola Banovic. 2016. To Replicate or Not to Replicate? *GetMobile: Mobile Comp. and Comm.,* 19, 4: 23–27. DOI:http://doi.acm.org/10.1145/2904337.2904346
5. Nikola Banovic, Christina Brant, Jennifer Mankoff, and Anind Dey. 2014. ProactiveTasks: The Short of Mobile Device Use Sessions. *Proc. MobileHCI '14*, 243–252. DOI:http://doi.acm.org/10.1145/2628363.2628380
6. Louise Barkhuus and Valerie E. Polichar. 2011. Empowerment through seamfulness: smart phones in everyday life. *Personal Ubiquitous Computing,* 15, 6: 629–639. DOI:http://dx.doi.org/10.1007/s00779-010-0342-4
7. Dror Ben-Zeev, Stephen M. Schueller, Mark Begale, Jennifer Duffecy, John M. Kane, and David C. Mohr. 2015. Strategies for mHealth Research: Lessons from 3 Mobile Intervention Studies. *Administration and Policy in Mental Health and Mental Health Services Research,* 42, 2: 157–167. DOI:https://doi.org/10.1007/s10488-014-0556-2
8. Niels van Berkel, Chu Luo, Theodoros Anagnostopoulos, Denzil Ferreira, Jorge Goncalves, Simo Hosio, and Vassilis Kostakos. 2016. A Systematic Assessment of Smartphone Usage Gaps. *Proc. CHI '16*, 4711–4721. DOI:http://doi.acm.org/10.1145/2858036.2858348
9. Matthias Böhmer, Brent Hecht, Johannes Schoning, Antonio Kruger, and Gernot Bauer. 2011. Falling Asleep with Angry Birds, Facebook and Kindle: A Large Scale Study on Mobile Application Usage. *Proc. MobileHCI '11*, 47–56. DOI:http://doi.acm.org/10.1145/2037373.2037383
10. Barry Brown, Stuart Reeves, and Scott Sherwood. 2011. Into the Wild: Challenges and Opportunities for Field Trial Methods. *Proc. CHI '11*, 1657–1666. DOI:http://doi.acm.org/10.1145/1978942.1979185
11. M. Butler. 2011. Android: Changing the Mobile Landscape. *IEEE Pervasive Computing,* 10, 1: 4-7.
12. Juan Pablo Carrascal and Karen Church. 2015. An In-Situ Study of Mobile App & Mobile Search Interactions. *Proc. CHI '15*, 2739–2748. DOI:http://doi.acm.org/10.1145/2702123.2702486
13. Karen Church, Denzil Ferreira, Nikola Banovic, and Kent Lyons. 2015. Understanding the Challenges of Mobile Phone Usage Data. *Proc. MobileHCI '15*, 504–514. DOI:http://doi.acm.org/10.1145/2785830.2785891
14. comScore. 2017. The 2017 U.S. Mobile App Report. Retrieved Sept 2017 from https://www.comscore.com/Insights/Presentations-and-Whitepapers/2017/The-2017-US-Mobile-App-Report
15. Chris Drummond. 2009. Replicability Is Not Reproducibility: Nor Is It Good Science. *Proceedings of Evaluation Methods for Machine Learning Workshop at the 26th ICML*.
16. Nathan Eagle and Alex Pentland. 2006. Reality mining: sensing complex social systems. *Personal and ubiquitous computing,* 10, 4: 255–268..
17. Denzil Ferreira, Jorge Goncalves, Vassilis Kostakos, Louise Barkhuus, and Anind K. Dey. 2014. Contextual Experience Sampling of Mobile Application Micro-usage. *Proc. MobileHCI '14*, 91–100. DOI:http://doi.acm.org/10.1145/2628363.2628367
18. Denzil Ferreira, Vassilis Kostakos, and Anind K. Dey. 2012. Lessons Learned from Large-Scale User Studies: Using Android Market As a Source of Data. *Int. J. Mob. Hum. Comput. Interact.,* 4, 3: 28–43. DOI:http://dx.doi.org/10.4018/jmhci.2012070102
19. Financial Times. 2010. Google's Android mobiles overtake global iPhone sales. Retrieved July 2017 from https://www.ft.com/content/77ed3ddc-a63f-11df-8767-00144feabdc0?mhq5j=e3
20. Forbes. 2014. What Kind Of Person Prefers An iPhone? Retrieved Jan 2018 from https://www.forbes.com/sites/toddhixon/2014/04/10/what-kind-of-person-prefers-an-iphone/#4f1a0be7d1b0
21. Forbes. 2017. Surprise: Google Reveals iOS Market Share Is 65% to 230% Bigger Than We Thought. Retrieved July 2017 from https://www.forbes.com/sites/johnkoetsier/2017/05/18/surprise-google-reveals-apples-ios-market-share-is-65-to-230-bigger-than-we-thought/#692389945890
22. Fortune. 2014. Apple's users spend 4X as much as Google's. Retrieved December 2017 from http://fortune.com/2014/06/27/apples-users-spend-4x-as-much-as-googles/
23. Andrea Girardello and Florian Michahelles. 2010. AppAware: which mobile applications are hot? *MobileHCI '10* 431–434.

24. The Guardian. 2014. Apps more popular than the mobile web, data shows . Retrieved July 2017 from https://www.theguardian.com/technology/appsblog/2014/apr/02/apps-more-popular-than-the-mobile-web-data-shows
25. Alina Hang, Alexander De Luca, Jonas Hartmann, and Heinrich Hussmann. 2013. Oh app, where art thou?: on app launching habits of smartphone users. *Proc. MobileHCI '13*, 392–395. DOI:http://doi.acm.org/10.1145/2493190.2493219
26. Jingjing Huangfu, Jian Cao, and Chenyang Liu. 2015. A Context-Aware Usage Prediction Approach for Smartphone Applications. *Advances in Services Computing. Lecture Notes in Computer Science* 9464, 3-16 2015.
27. IDC. 2017. Smartphone OS Market Share, 2017 Q1. Retrieved July 2017 from http://www.idc.com/promo/smartphone-market-share/os
28. The Independent. 2016. Twitter changes itself from a social network into a news app. Retrieved Sept 2017 from http://www.independent.co.uk/life-style/gadgets-and-tech/news/twitter-changes-itself-from-a-social-network-into-a-news-app-a7006696.html
29. Uichin Lee, Joonwon Lee, Minsam Ko, Changhun Lee, Yuhwan Kim, Subin Yang, Koji Yatani, Gahgene Gweon, Kyong-Mee Chung, and Junehwa Song. 2014. Hooked on Smartphones: An Exploratory Study on Smartphone Overuse Among College Students. *Proc. CHI '14*, 2327–2336. DOI:http://doi.acm.org/10.1145/2556288.2557366
30. Carolynne Lord, Mike Hazas, Adrian K. Clear, Oliver Bates, Rosalind Whittam, Janine Morley, and Adrian Friday. 2015. Demand in My Pocket: Mobile Devices and the Data Connectivity Marshalled in Support of Everyday Practice, 2729-2738.
31. Donald McMillan, Moira McGregor, and Barry Brown. 2015. From in the Wild to in Vivo: Video Analysis of Mobile Device Use. *Proc. MobileHCI '15*, 494–503. DOI:http://doi.acm.org/10.1145/2785830.2785883
32. Donald McMillan, Alistair Morrison, and Matthew Chalmers. 2011. A Comparison of Distribution Channels for Large-Scale Deployments of iOS Applications. *IJMHCI,* 3, 4: 1–17.
33. Donald McMillan, Alistair Morrison, and Matthew Chalmers. 2013. Categorised ethical guidelines for large scale mobile HCI. *Proc. CHI '13*, 1853–1862. DOI:http://doi.acm.org/10.1145/2466110.2466245
34. Moon Technolabs Pvt Ltd. 2017. Apple Vs Android – A comparative study 2017. Retrieved Jan 2018 from https://www.moontechnolabs.com/apple-vs-android-comparative-study-2017/
35. Ofcom. 2015. The Communications Market Report: United Kingdom. Retrieved July 2017 from https://www.ofcom.org.uk/research-and-data/multi-sector-research/cmr/cmr15/uk
36. PCWorld. 2010. 5 Reasons to Jailbreak Your iPhone - and 5 Reasons Not. Retrieved December 2017 from https://www.pcworld.com/article/202441/5_Reasons_to_Jailbreak_Your_iPhone_and_5_Reasons_Not_To.html
37. Benjamin Poppinga, Henriette Cramer, Matthias Bohmer, Alistair Morrison, Frank Bentley, Niels Henze, Mattias Rost, and Florian Michahelles. 2012. Research in the large 3.0: app stores, wide distribution, and big data in MobileHCI research. *Proc. MobileHCI '12*, 241–244. DOI:http://doi.acm.org/10.1145/2371664.2371724
38. John Rooksby, Parvin Asadzadeh, Mattias Rost, Alistair Morrison, and Matthew Chalmers. 2016. Personal Tracking of Screen Time on Digital Devices. *Proc. CHI '16*, 284–296. DOI:http://doi.acm.org/10.1145/2858036.2858055
39. John Rooksby, Parvin Asadzadeh, Alistair Morrison, Claire McCallum, Cindy Gray, and Matthew Chalmers. 2016. Implementing Ethics for a Mobile App Deployment. *Proc. OzCHI '16*, 406–415. DOI:http://doi.acm.org/10.1145/3010915.3010919
40. Alireza Sahami, Niels Henze, Tilman Dingler, Kai Kunze, and Albrecht Schmidt. 2013. Upright or sideways?: analysis of smartphone postures in the wild. *Proc. MobileHCI '13*, 362–371. DOI:http://doi.acm.org/10.1145/2493190.2493230
41. Scientific American. 2015. Apple's First 5 Health ResearchKit Apps in Brief. Retrieved Feb 2017 from https://www.scientificamerican.com/article/pogue-apples-first-5-health-researchkit-apps-in-brief/
42. Choonsung Shin, Jin-Hyuk Hong, and Anind K. Dey. 2012. Understanding and Prediction of Mobile Application Usage for Smart Phones. *Proc. UbiComp '12*, 173–182. DOI:http://doi.acm.org/10.1145/2370216.2370243
43. Tech Insider. 2013. The Latest Jailbreak Statistics Are Jaw-Dropping. Retrieved Feb 2017 from http://www.businessinsider.com/jailbreak-statistics-2013-3?IR=T
44. The Telegraph. 2017. Google's Android more popular than Windows for first time . Retrieved July 2017 from http://www.telegraph.co.uk/technology/2017/04/03/googles-android-popular-windows-first-time/
45. Seppo Virtanen, Mattias Rost, Alistair Morrison, Matthew Chalmers, and Mark Girolami. 2016. Uncovering smartphone usage patterns with multi-view mixed membership models. *Stat,* 5, 1: 57–69. DOI:http://dx.doi.org/10.1002/sta4.103

46. Pascal Welke, Ionut Andone, Konrad Blaszkiewicz, and Alexander Markowetz. 2016. Differentiating Smartphone Users by App Usage. *Proc. UbiComp '16*, 519–523. DOI:http://doi.acm.org/10.1145/2971648.2971707

47. Max L. Wilson, Wendy Mackay, Ed Chi, Michael Bernstein, Dan Russell, and Harold Thimbleby. 2011. RepliCHI - CHI Should Be Replicating and Validating Results More: Discuss. *Proc. CHI EA '11*, 463–466. DOI:http://doi.acm.org/10.1145/1979742.1979491

48. Qiang Xu, Jeffrey Erman, Alexandre Gerber, Zhuoqing Mao, Jeffrey Pang, and Shobha Venkataraman. 2011. Identifying diverse usage behaviors of smartphone apps. *Proc. IMC '11*, 329–344. DOI:http://doi.acm.org/10.1145/2068816.2068847

49. Jiangchuan Zheng and Lionel M Ni. 2012. An unsupervised framework for sensing individual and cluster behavior patterns from human mobile data. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* 153–162.